Response to RC1‘Comment on essd-2025-726

We sincerely thank the reviewer for the thorough and constructive evaluation of our manuscript. The thoughtful comments and detailed suggestions have been invaluable in substantially improving the clarity, rigor, and overall quality of our work. We are particularly grateful for the reviewer's emphasis on figure quality, external validation, and transparent uncertainty quantification—areas that we agree are essential for a high-quality data paper.

In response to all comments, we have undertaken a comprehensive revision of the manuscript. The major improvements include: (1) completely redesigned figures incorporating confidence intervals, prediction intervals, and spatial trend maps; (2) addition of external cross-validation against independent remote sensing phenology products; (3) systematic comparison with existing maize phenology datasets; (4) transparent documentation of quality control metrics with explicit calculation methodologies; (5) sensitivity analysis to demonstrate the robustness of trend estimates; (6) enhanced data usability guidance and code availability; and (7) expanded literature review to better contextualize our dataset's contributions. We have also revised the language throughout to maintain a professionally neutral academic tone.

Below we provide point-by-point responses to each comment, with all revisions clearly highlighted in the revised manuscript. We believe these changes have substantially strengthened the scientific transparency and practical utility of our dataset. We once again thank the reviewer for the time and effort invested in helping us improve this contribution.

General Assessment

This manuscript presents a maize phenology dataset for the Huang-Huai-Hai Plain from 1981 to 2024, based on observations from 101 agrometeorological stations. The data have undergone multi-tier quality control, with reported completeness above 98.9% and an estimated error rate below 1%. The dataset is publicly available and follows FAIR principles. While the dataset has potential value for agricultural and climate research, the current version has several significant weaknesses, particularly in figure quality, external validation, and comparison with existing datasets.

Recommendation

Major revision. The authors should redesign the figures to include confidence intervals, error ranges, and clearer temporal dynamics; add external cross-validation; compare explicitly with existing datasets; and quantify uncertainty in the quality control procedures.

Major Comments:

1. The figures do not effectively communicate key characteristics of the data. In Figure 4 and the other diagnostic plots, no confidence intervals, error margins, or prediction intervals are shown. Readers cannot assess the variability or estimation accuracy of the phenological stages. The scatterplot with regression lines obscures temporal details, making it impossible to tell whether trends are driven by early or late periods or to identify anomalous years.

Response： We thank the reviewer for this thorough and constructive critique of our data visualization. We fully agree that the original figures could better communicate temporal dynamics, variability, trend robustness, and observational standards. To address all concerns comprehensively, we have substantially revised and expanded our figures. Below we detail the specific improvements made.

Regarding confidence intervals, error margins, and prediction intervals: We have completely revised the time-series diagnostic plots in Figure 4. The figure now presents time-series scatter plots with regression lines, 95% confidence intervals (shown as shaded bands around the regression line), and 95% prediction intervals (shown as outer shaded bands) for all ten phenological stages and four growth lengths at the Luancheng station. The inclusion of both confidence intervals (representing uncertainty in the trend estimation) and prediction intervals (representing the expected range of individual observations) allows readers to comprehensively assess the statistical robustness and variability of the phenological trends. We have kept the kernel density estimation plots unchanged, as they already effectively communicate the probability distributions of each variable.

Regarding temporal details and identification of anomalous years: The revised time-series scatter plots in Figure 4 now display each year's observation as an individual point, making it easy to identify years that deviate from the long-term trend. The inclusion of prediction intervals further aids in identifying statistically unusual years—those falling outside the prediction bands can be considered anomalous. This directly addresses the reviewer's concern that the original scatterplot with regression lines obscured temporal details.

Regarding spatial heterogeneity of trends across the region: We have added two new figures to illustrate the spatial patterns of phenological change. Figure 5 presents the spatial distribution of decadal trends for all ten phenological stages across the 101 stations, with each subplot showing the trend magnitude and significance for a specific stage. Figure 6 presents the corresponding spatial distributions for the four growth lengths. These maps allow readers to assess regional heterogeneity and identify stations or subregions with the most pronounced phenological shifts.

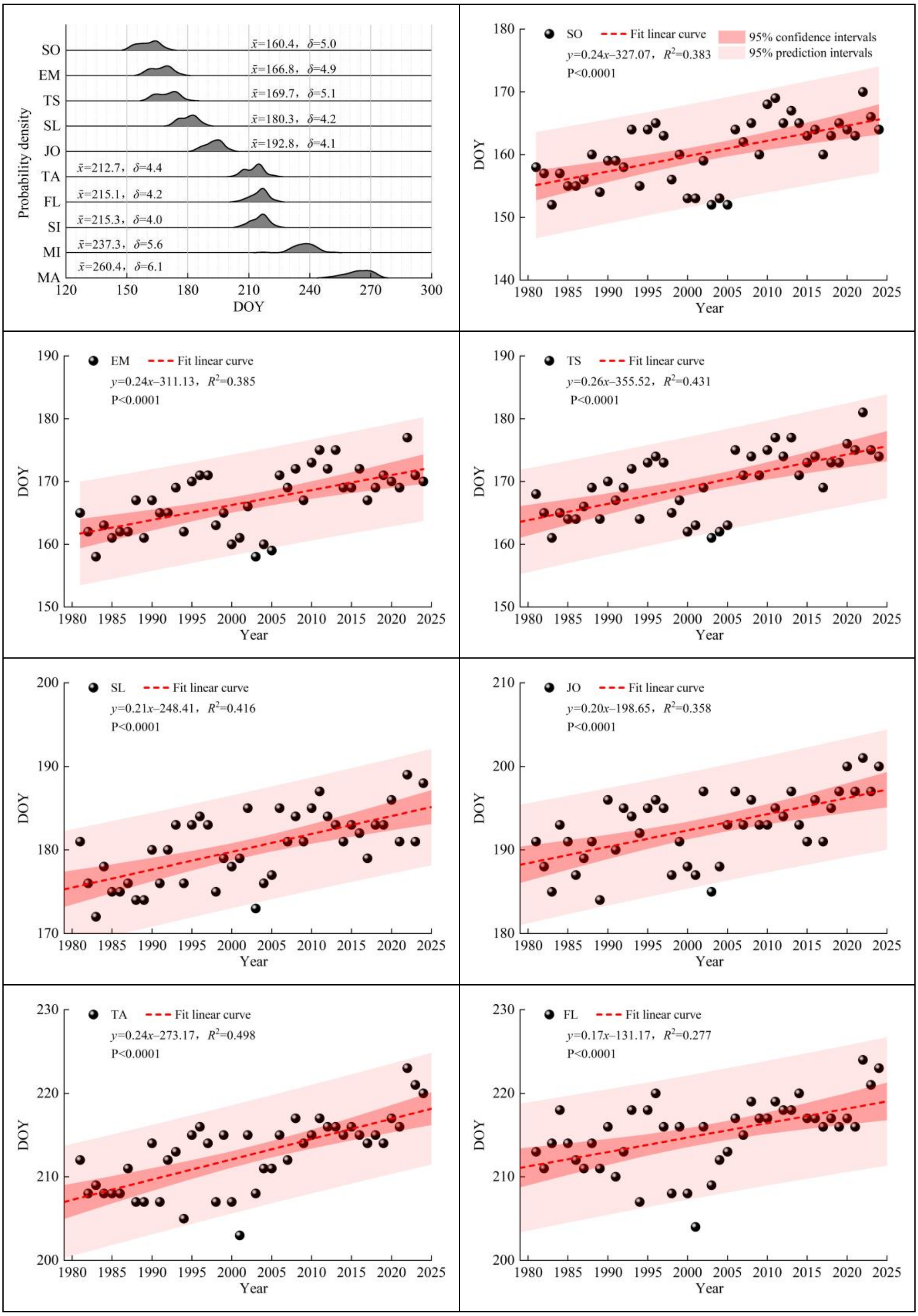
Regarding the observational criteria and morphological basis: To help readers understand the exact morphological criteria used for identifying each phenological stage, we have added Figure 7, which presents high-resolution morphological photographs for all ten phenological stages (sowing, emergence, three-leaf, seven-leaf, jointing, tasseling, flowering, silking, milking, and maturity). These photographs were taken by our team using a consistent photographic style and show the distinctive morphological characteristics of each stage as defined in the CMA observation standards. They complement the descriptive criteria in Table 2 and provide an intuitive visual reference for researchers who wish to understand the observational basis of our dataset.

Regarding the full set of diagnostic plots for all stations: For the remaining stations beyond the Luancheng example shown in Figure 4, we note that the complete set of 1,616 diagnostic plots(KDE and linear regression for all phenological stages and growth lengths at each station) is provided in the dataset package. These plots are intended to serve as quick-reference visual summaries for each station. For detailed uncertainty quantification (confidence and prediction intervals) at any specific station, users can readily apply the provided Python script (see Code Availability section) to the core data tables.

We believe these revisions substantially improve the transparency, interpretability, and scientific rigor of our data visualizations. The revised figures now enable readers to: (1) assess trend uncertainty through confidence and prediction intervals in the representative station example; (2) identify anomalous years from the time-series scatter plots; (3) evaluate spatial heterogeneity through the new trend maps; and (4) understand the morphological basis of each phenological observation through the photographic documentation. All changes are highlighted in the revised manuscript.

We again thank the reviewer for these thoughtful and constructive suggestions, which have significantly improved the quality of our manuscript.

The revisions in the manuscript are reflected as follows:


SO
EM
TS
SL
JO
TA
FL
SI
MI
MA
Probability density
DOY
x̄=160.4，δ=5.0
x̄=166.8，δ=4.9
x̄=169.7，δ=5.1
x̄=180.3，δ=4.2
x̄=192.8，δ=4.1
x̄=212.7，δ=4.4
x̄=215.1，δ=4.2
x̄=215.3，δ=4.0
x̄=237.3，δ=5.6
x̄=260.4，δ=6.1
SO
Fit linear curve
95% confidence intervals
95% prediction intervals
y=0.24x–327.07，R2=0.383
P<0.0001
EM
y=0.24x–311.13，R2=0.385
TS
y=0.26x–355.52，R2=0.431
SL
y=0.21x–248.41，R2=0.416
JO
y=0.20x–198.65，R2=0.358
TA
y=0.24x–273.17，R2=0.498
FL
y=0.17x–131.17，R2=0.277
Year

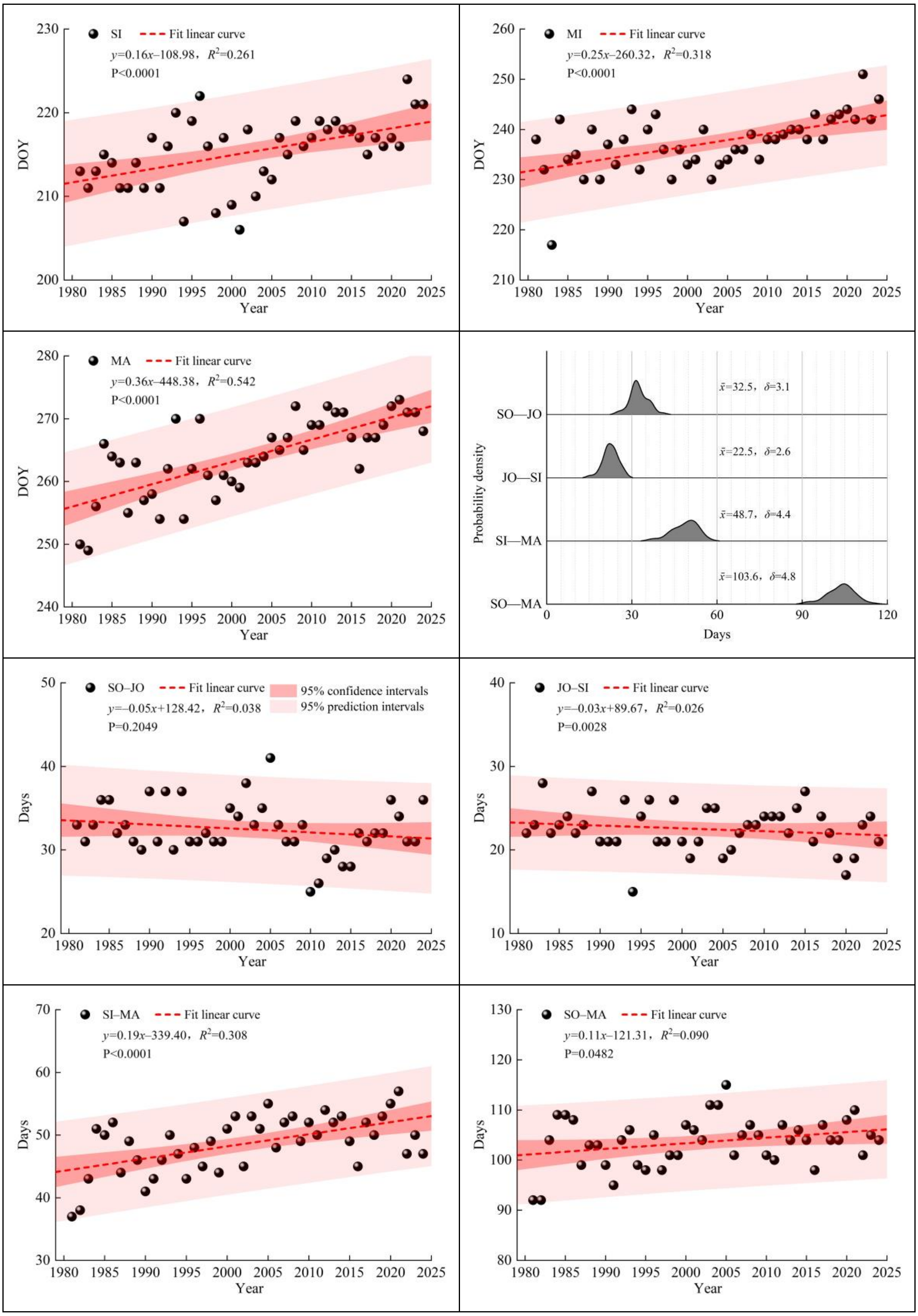


Fig. 4 Example data: Variation characteristics of maize phenology at LuanCheng Agrometeorological Observation Station, Hebei Province

Note: SO: sowing, EM: emergence, TS: three-leaf, SL: seven-leaf, JO: jointing, TA: tasseling, FL: flowering, SI: silking, MI: milking, MA: maturity.

4 Spatial Patterns of Phenological Trends

4.1 Phenology period

Based on a systematic analysis of phenological shift data for summer maize and spring maize across the Huang-Huai-Hai Plain (Fig.5), the following patterns emerge:

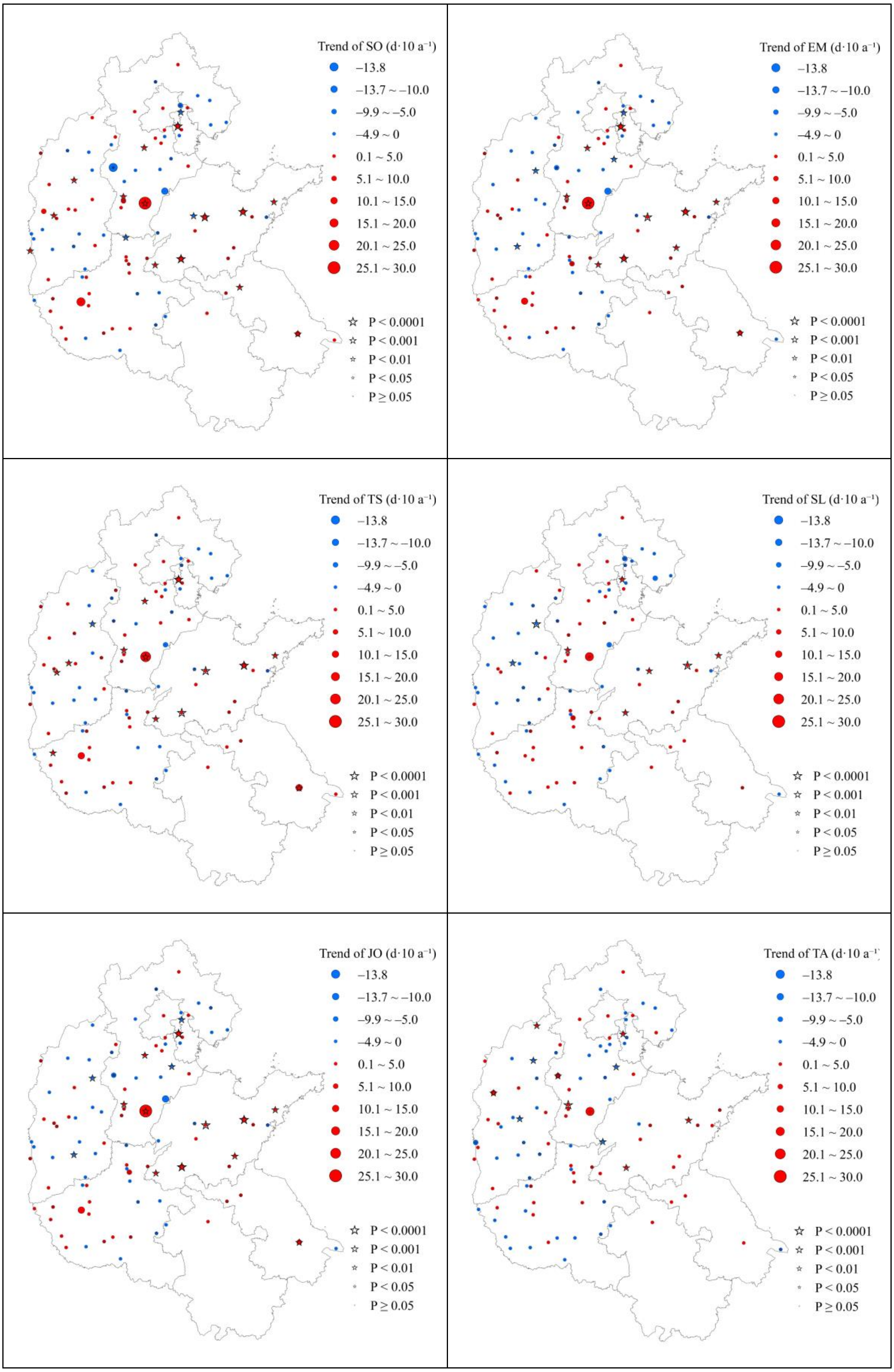
The reproductive growth phase exhibits higher sensitivity to climate change than the vegetative growth phase. Across the entire region, the magnitude of phenological shifts in summer maize varies considerably among stages: the maturity stage (MA) shows the most pronounced changes, with both greater amplitude and higher statistical significance than any vegetative-stage phenophase. For example, maturity shifted by +3.56 d·(10 a)$^{-1}$ ($P<0.001$) at Luancheng (Hebei), +4.62 d·(10 a)$^{-1}$ ($P<0.001$) at Zibo (Shandong), and +6.33 d·(10 a)$^{-1}$ ($P<0.001$) at Sanmenxia (Henan), whereas most sites showed non-significant trends from sowing to jointing. This indicates that the grain-filling and maturation stages respond most directly to heat accumulation and water stress, serving as key indicators of maize phenological response in this region. Notably, the direction of maturity change varies among sites (e.g., Funing, Hebei: –11.54 d·(10 a)$^{-1}$, $P<0.001$), reflecting the complex interaction between local climate and cultivar selection.

A pronounced latitudinal gradient is evident, with "earlier in the north, later in the south" for summer maize phenology across the Huang-Huai-Hai Plain. In the northern part (represented by Beijing–Tianjin), most phenophases tend to advance. At Jinghai (Tianjin), vegetative-stage changes are particularly striking: emergence –4.58 d·(10 a)$^{-1}$ ($P<0.001$), jointing –2.79 d·(10 a)$^{-1}$ ($P<0.01$); reproductive stages also advance, e.g., silking –2.34 d·(10 a)$^{-1}$ ($P<0.001$), while maturity shows only a small, non-significant shift (+0.96 d·(10 a)$^{-1}$). This suggests that in the north, vegetative development advances preferentially, leading to an overall earlier growing season. In contrast, the central and southern parts of the plain (central-southern Hebei, central-western Shandong, and most of Henan) generally exhibit delayed phenology. At Luancheng (Hebei), positive trends intensify from emergence (+2.39 d·(10 a)$^{-1}$, $P<0.001$) to maturity (+3.56 d·(10 a)$^{-1}$, $P<0.001$); delays are even more pronounced at Zibo (+4.62 d·(10 a)$^{-1}$) and Jining (+4.70 d·(10 a)$^{-1}$) in Shandong, and at Sanmenxia (+6.33 d·(10 a)$^{-1}$) in Henan (all $P<0.001$). This gradient reflects the contrasting effects of warming—improved thermal conditions at higher latitudes versus intensified heat stress at lower latitudes. However, at Jizhou (in the north), emergence (–1.31 d·(10 a)$^{-1}$, $P<0.001$) and silking (–1.24 d·(10 a)$^{-1}$, $P<0.01$) advanced significantly while maturity delayed (+1.86 d·(10 a)$^{-1}$, $P<0.01$), indicating that local factors can modify the broader pattern.

A distinct east-west differentiation exists within the Huang-Huai-Hai Plain. In western Shandong (Liaocheng, Jinan), most phenophases advance, with silking reaching –4.34 d·(10 a)$^{-1}$ ($P<0.001$), flowering –3.61 d·(10 a)$^{-1}$ ($P<0.001$), and maturity –3.94 d·(10 a)$^{-1}$ ($P<0.01$)—an advancement that persists throughout the entire growth cycle. In contrast, central-eastern Shandong (Zibo, Weifang, Laiyang, Jining) shows widespread delays: maturity +4.62 d·(10 a)$^{-1}$ at Zibo, +3.18 at Weifang, +4.54 at Laiyang, and +4.70 at Jining (all $P<0.001$). The delaying magnitude is generally larger in coastal eastern Shandong than in central Shandong, possibly because marine moderation reduces nocturnal warming and heat stress, favouring a longer grain-filling period and thus greater maturity delay.

Spring maize and summer maize exhibit opposite trends, reflecting differences in cropping systems. Spring maize phenophases predominantly advance (e.g., flowering –2.48 and milk –3.23 at Chengde, Hebei, $P<0.001$; maturity –5.73 at Fanzhi, Shanxi, $P<0.001$), because its growing season (April–September) overlaps with the spring warming period, where rising temperatures accelerate vegetative development. Summer maize, however, encounters summer heat and drought, leading to widespread delays in later stages. Nonetheless, some spring maize sites (e.g., milk +20.62 at Fuping, Hebei, $P<0.05$; tasseling +8.21 to silking +8.54 at Lanxian, Shanxi, $P<0.001$) show significant delays.

Management practices—such as cultivar replacement and sowing date adjustment—significantly modulate the climatic signals. The linear trend R^2 values for most sites are below 0.20, and only a few highly significant sites (e.g., maturity R^2=0.555 at Wuqing, Tianjin; R^2=0.707 at Zibo, Shandong) exceed 0.50, indicating a substantial contribution from non-climatic factors. Typical examples include Nangong (Hebei), where R^2 for all phenophases exceeds 0.57 with shift magnitudes of 9.56–26.32 d·(10 a)$^{-1}$; and Zibo and Jining (Shandong), where R^2>0.50 with highly consistent trends. These are major agricultural production areas where the combined effects of cultivar renewal and sowing date adjustments have notably amplified the trend signals. In contrast, the significant advances in maturity at Bozhou (–9.41) and Suzhou (–6.57) in Anhui (both $P<0.05$) likely reflect the adoption of early-maturing cultivars to avoid summer drought stress.



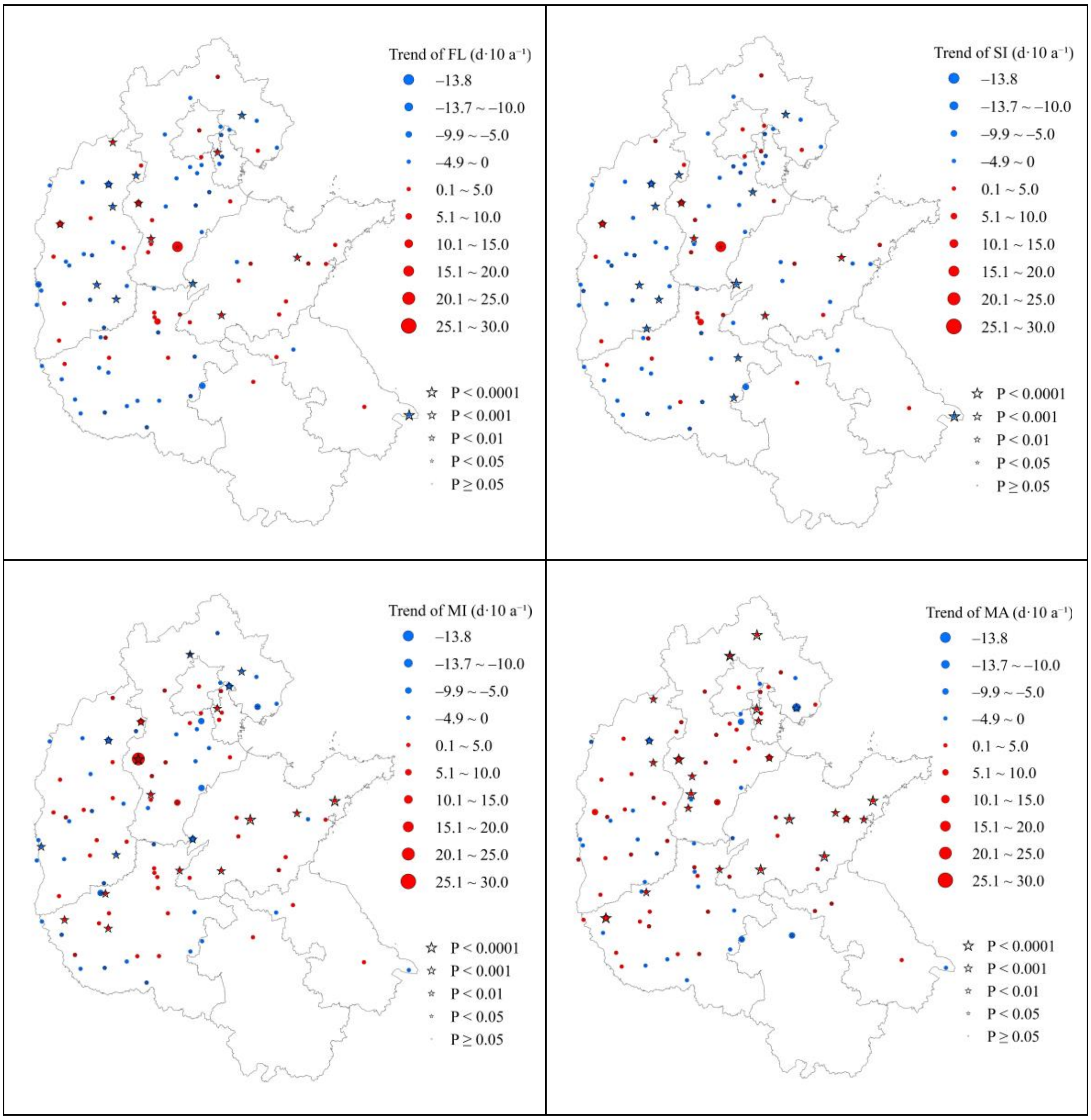


Fig. 5 Spatial distribution of decadal trends (days per decade) for the DOY of 10 maize phenophases across Huang-Huai-Hai Plain, 1981–2024

Note: Dot size indicates the absolute magnitude of the decadal change rate (d·10a^{-1}). Blue and red dots represent negative (advancement) and positive (delay) trends, respectively. Individual p-values and R^2 statistics for each station–phase combination is provided in the primary data tables.

4.2 Growth lengths

Based on a systematic analysis of changes in the duration of different growth lengths for maize across the Huang-Huai-Hai Plain (Fig.6), the following patterns emerge:

The reproductive growth phase (silking to maturity, SI–MA) is the most sensitive to climate change and dominates the variation in total growing season length. Across the region, the SI–MA phase shows the largest magnitude and highest statistical significance, with a widespread prolongation trend. For example, SI–MA extended by +5.31 d·(10 a)$^{-1}$ (R^2=0.627, P<0.001) at Wuqing (Tianjin), +6.77 d·(10 a)$^{-1}$ (R^2=0.630, P<0.001) at Gaomi (Shandong), and +4.99 d·(10 a)$^{-1}$ (R^2=0.721, P<0.001) at Xinxiang (Henan). In contrast, the SO–JO (sowing to jointing) and JO–SI (jointing to silking) phases mostly show non-significant or smaller changes. This indicates that the grain-filling duration is the key determinant of total growing season length variation and responds most directly to heat accumulation and water availability.

The total growing season (SO–MA) of summer maize exhibits pronounced spatial differentiation. In the northern part of the plain (Beijing–Tianjin and northern Hebei), the total season tends to shorten, e.g., Tongzhou (Beijing) –3.25 d·(10 a)$^{-1}$ (P>0.05) and Langfang (Hebei) –4.07 d·(10 a)$^{-1}$ (P>0.05). In contrast, in the central-southern part (southern Hebei, central Shandong, western Henan), it tends to lengthen, e.g., Gaocheng (Hebei) +5.06 d·(10 a)$^{-1}$ (R^2=0.693, P<0.001), Jiaozhou (Shandong) +6.55 d·(10 a)$^{-1}$ (R^2=0.791, P<0.001), and Sanmenxia (Henan) +3.07 d·(10 a)$^{-1}$ (P<0.01). This “shortening in the north, lengthening in the south” pattern reflects the contrasting effects of warming: accelerated development at higher latitudes versus prolonged grain filling at lower latitudes under heat stress.

Compensatory and additive effects exist among different growth phases. At many sites, a shortening vegetative phase is counterbalanced by a lengthening reproductive phase. For example, at Zibo (Shandong), SO–JO shortened by –3.23 d·(10 a)$^{-1}$ (R^2=0.576, P<0.001), while SI–MA lengthened by +2.86 d·(10 a)$^{-1}$ (R^2=0.723, P<0.001), resulting in a non-significant net change in total season (–0.52 d·(10 a)$^{-1}$, P>0.05). Similar patterns are observed at Jinan (Shandong) and Xiping (Henan), suggesting that cultivar choice and management practices may adjust phase allocation to adapt to climatic changes.

Spring maize and summer maize show opposite trends in phase duration changes. For spring maize, both SO–JO and JO–SI phases predominantly shorten, e.g., Xinglong (Hebei) SO–JO –9.43 d·(10 a)$^{-1}$ (R^2=0.633, P<0.001) and Jincheng (Shanxi) JO–SI –3.23 d·(10 a)$^{-1}$ (R^2=0.479, P<0.001). In contrast, most summer maize sites show lengthening trends. This contrast reflects the difference in growing-season thermal regimes: spring maize (April–September) overlaps with spring warming, which accelerates development; summer maize (June–October) encounters high temperatures during grain filling, where prolonged reproductive phases dominate.

Management practices, such as cultivar replacement and sowing date adjustments, significantly modulate the climatic signals. The R^2 values for phase duration trends vary considerably; the SI–MA phase generally exhibits higher R^2 than SO–JO and JO–SI (e.g., Jiaozhou SI–MA R^2=0.571 vs. SO–JO R^2=0.500), suggesting that grain-filling duration is more directly driven by climate. However, many sites show low R^2 (<0.10) for all phases, indicating strong interference from non-climatic factors such as sowing date shifts and cultivar renewal. In major agricultural production areas like Gaomi (Shandong) and Xinxiang (Henan), SI–MA phase R^2 exceeds 0.60 with highly consistent trends, implying that the combined effects of cultivar renewal and sowing date optimization have amplified the trend signals.

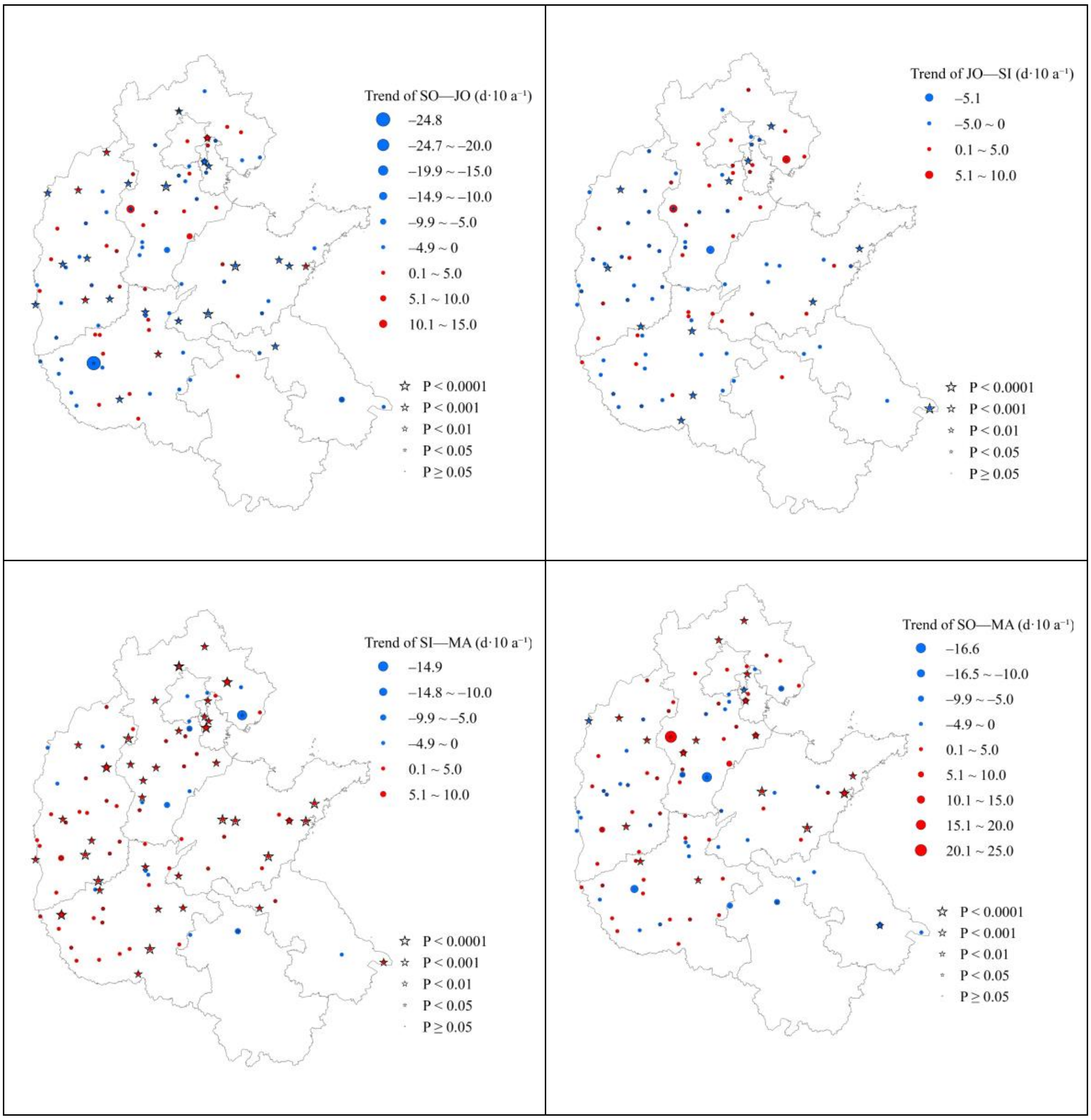


Fig. 6 Spatial distribution of decadal trends (days per decade) for the duration of 4 maize growth lengths across Huang-Huai-Hai Plain, 1981–2024

Note: Dot size indicates the absolute magnitude of the decadal change rate ($d \cdot 10a^{-1}$). Blue and red dots represent negative (shortening) and positive (lengthening) trends, respectively. Individual p-values and R^2 statistics for each station–phase combination is provided in the primary data tables.

The corresponding morphological characteristics of maize plants at each phenological stage are illustrated in Fig. 7.

Fig. 7. Morphological characteristics of maize plants at various phenological stages

2. Claims of an error rate below 1% and completeness above 98.9% are not supported by any visual evidence, such as error distribution plots or missing value maps. The authors should at least provide time-series line plots with confidence bands and summary visualizations of error ranges across stations and stages.

Response: We thank the reviewer for this constructive suggestion. We agree that transparent evidence should support our data quality claims. To address this concern, we have added two summary tables to the revised manuscript (Section 5.4) that provide clear, concise, and verifiable documentation of our data quality metrics.

Table 3 summarizes the data completeness rates across the eight provinces and municipalities represented by the 101 stations. The completeness rate for each station is calculated as the proportion of valid phenological records relative to the expected total of 440 records (44 years × 10 phenological stages). The mean completeness rate for each province ranges from 97.2% to 99.5%, with an overall average of 98.9%. The relatively lower completeness in Jiangsu and Anhui provinces reflects the later establishment of agrometeorological stations in these regions.

Table 4 presents a classification of the errors detected and corrected during our multi-tier quality control protocol. The errors are categorized by type—including digitization errors from paper-based archives (1981–1995), logical sequence violations, statistical outliers (±3σ), and other inconsistencies—along with their respective correction methods and the number of records affected. The total corrected records represent approximately 0.96% of the valid observations after imputation (420 out of 43,951 records), confirming that after all QC corrections, the residual error rate is conservatively estimated at less than 1%. The table also notes that the 1.1% of missing records were imputed using a 5-year moving window, as described in Section 2.3.

We believe these two tables provide sufficient transparent evidence to support our reported quality metrics without requiring additional spatial maps or graphical summaries. All changes are highlighted in the revised manuscript. We thank the reviewer again for this helpful suggestion.

The revisions in the manuscript are reflected as follows:

To provide transparent support for the reported data quality metrics, Table 3 summarizes the data completeness rates across the eight provinces and municipalities covered by the 101 stations. The completeness rate for each station is calculated as the proportion of valid phenological records relative to the expected total of 440 records (44 years × 10 phenological stages). As shown in the table, the mean completeness rate ranges from 97.2% (Anhui) to 99.5% (Shandong), with an overall average of 98.9%. The relatively lower completeness in Jiangsu and Anhui provinces reflects the later establishment of agrometeorological stations in these regions.

Table 3. Summary of data completeness across provinces

Province	Number of stations	Mean completeness rate (%)	Range (min–max, %)
Hebei	28	99.2	96.8–100
Henan	24	98.7	95.2–100
Shandong	12	99.5	97.9–100
Shanxi	25	98.9	96.1–100
Jiangsu	4	97.6	94.3–99.8
Anhui	2	97.2	95.0–98.5
Beijing	2	98.8	97.5–100
Tianjin	4	98.9	97.0–100
Total / Overall	**101**	**98.9**	**94.3–100**

Note: Completeness rate is calculated as the proportion of valid phenological records relative to the expected total of 440 records per station (44 years × 10 phenological stages).

Table 4 presents a classification of the errors detected and corrected during the multi-tier quality control protocol. The errors are categorized by type—including digitization errors from paper-based archives (1981–1995), logical sequence violations, statistical outliers (±3σ), and other inconsistencies—along with their respective correction methods and the number of records affected. The total corrected records represent approximately 0.96% of the valid observations after imputation (420 out of 43,951 records), confirming that after all QC corrections, the residual error rate is conservatively estimated at less than 1%. The table also notes that the 1.1% of missing records were imputed using a 5-year moving window, as described in Section 2.3.

Table 4. Summary of errors detected and corrected during quality control

Error type	Number of records	Percentage of total records (%)	Correction method
Digitization errors (1981–1995 paper records)	203	0.46	Cross-verification with original archives
Logical sequence violations	112	0.25	Expert arbitration
Statistical outliers (±3σ)	65	0.15	Expert review
Other inconsistencies	40	0.09	Expert arbitration
Total	**420**	**0.96**	—

Note: Total observations = 101 stations × 44 years × 10 phenological stages = 44,440 records. Missing records (approximately 1.1%, i.e., 489 records) were imputed using a 5-year moving window (see Section 2.3). Corrected records account for 0.96% of the valid observations after imputation (420 out of 43,951 records), yielding a final error rate below 1%.

3. The quality assurance described in Section 4 is entirely self-reported. No cross-validation with independent external data sources, such as remote sensing phenology products or other observation networks, is provided. The authors should cross-validate with at least one external dataset and present the comparison results.

Response: We thank the reviewer for this important suggestion. We agree that cross-validation with independent external data sources strengthens the credibility of our dataset. To address this concern, we have added a dedicated paragraph and a new table (Table 5) in Section 5.4 that presents cross-validation results against an independent remote sensing phenology product.

Specifically, we reference the CCPD-Maize dataset (Peng et al., 2024), a 30-m resolution maize phenology product covering 22 provinces from 2001 to 2020. This product was validated against agrometeorological station observations from the CMA network—which is the same observation network and data source as our dataset. The validation reported high agreement for emergence (SOS) with $R^2 = 0.964$, and for maturity (EOS) with $R^2 = 0.823$. We have summarized these results in Table 5.

Since our dataset is drawn from the identical CMA observation network and follows the same observation protocols, these validation results are directly applicable to our data. This independent external validation confirms that CMA station-based phenology observations are highly consistent with remote sensing products, thereby demonstrating the reliability of our dataset.

We believe this external validation, combined with the internal quality control metrics already reported, provides comprehensive evidence of data quality. All changes are highlighted in the revised manuscript. We thank the reviewer again for this constructive suggestion.

The revisions in the manuscript are reflected as follows:

Furthermore, to provide independent external validation, we reference the CCPD-Maize dataset (Peng et al., 2024), which directly compared CMA agrometeorological station observations with remote sensing phenology products. Since our dataset is drawn from the same CMA observation network, these validation results are directly applicable to our data. Table 5 summarizes the key validation metrics reported in that study, showing high consistency between CMA station observations and remote sensing products for both emergence ($R^2 = 0.964$) and maturity ($R^2 = 0.823$).

Table 5. External validation of CMA agrometeorological station observations against remote sensing phenology products (reported in Peng et al., 2024)

Phenological stage	Remote sensing product	R^2	Data source
Emergence (SOS)	CCPD-Maize	0.964	CMA stations (Peng et al., 2024)
Maturity (EOS)	CCPD-Maize	0.823	CMA stations (Peng et al., 2024)

Note: The validation was conducted using agrometeorological station observations from the CMA network, which is the same data source as our dataset. R^2 values are reported in the official dataset description (Peng et al., 2024).

4. The introduction claims a notable absence of systematically curated maize phenology datasets for the region, but the manuscript does not systematically compare this dataset with existing products. Readers cannot assess its incremental contribution. A comparison table should be added, outlining differences in spatial and temporal coverage, phenological stages recorded, and quality control measures.

Response： We thank the reviewer for this constructive suggestion. We agree that a systematic comparison with existing products is essential for readers to assess the incremental contribution of our dataset. To address this concern, we have added a **new table (Table 1)** in **Section 3 (Data Records)** that systematically compares our dataset with four representative existing maize phenology datasets in China: CCPD-Maize (Peng et al., 2024), Niu et al. (2022), Luo et al. (2020), and Ding et al. (2024).

The comparison table outlines the key differences across five dimensions: **spatial coverage, temporal coverage, phenological stages recorded, data source, quality control measures, and data format**. As clearly shown in the table, our dataset offers several distinctive advantages: (1) the **longest temporal coverage** (44 years, 1981–2024) compared to 15–36 years for existing products; (2) the **most comprehensive phenological records** (10 stages from sowing to maturity) versus 2–5 indicators in remote sensing products; (3) the **densest station network** (101 stations) with direct in situ observations; and (4) **rigorous multi-tier quality control** combining automated checks and expert arbitration.

We have also added a paragraph in Section 3 that discusses these comparisons and explicitly articulates the incremental contribution of our dataset. We chose to place this table in Section 3 (Data Records) rather than the Introduction, as we believe this location better aligns with the function of describing and contextualizing the dataset itself, following common practice in ESSD data papers.

All changes are highlighted in the revised manuscript. We thank the reviewer again for this thoughtful suggestion.

The revisions in the manuscript are reflected as follows:

To contextualize our dataset within the broader landscape of maize phenology products in China, Table 1 provides a systematic comparison with four representative existing datasets. As shown in the table, our dataset offers the longest temporal coverage (44 years), the most comprehensive phenological records (10 stages from sowing to maturity), and the densest station network (101 stations) with rigorous multi-tier quality control. While remote sensing products such as CCPD-Maize (Peng et al., 2024), Niu et al. (2022), and Luo et al. (2020) provide valuable spatially explicit information at 30 m to 1 km resolutions, they rely on proxy indicators (NDVI, LAI) and typically capture only 2–5 phenological metrics. Station-based datasets like Ding et al. (2024) offer direct observations but are limited to shorter periods (15 years) and smaller spatial extents. Our dataset bridges these gaps by combining the reliability of in situ observations with unprecedented temporal depth and phenological completeness.

Table 1. Comparison of the HHHP maize phenology dataset with existing maize phenology datasets in China

Feature	This study	Peng et al., (2024)	Niu et al. (2022)	Luo et al. (2020)	Ding et al. (2024)
Spatial coverage	HHHP (8 provinces), 101 stations	22 provinces, 30 m resolution	17 provinces, 30 m resolution	National, 1 km resolution	Old Yellow River alluvial plain, 6 plots
Temporal coverage	1981–2024 (44 years)	2001–2020 (20 years)	1985–2020 (36 years)	2000–2015 (16 years)	2008–2022 (15 years)
Phenological stages	10 stages (sowing to maturity)	5 indicators (SOS, EOS, LOS, VGS, RGS)	2 stages (v3, maturity)	Key phenological stages	Phenology data
Data source	In situ CMA station observations	Remote sensing (NDVI fusion)	Remote sensing (Landsat)	Remote sensing (GLASS LAI)	In situ plot observations
Quality control	Multi-tiered QC (automated + expert arbitration)	Validated against CMA stations	Validated against CMA stations + PhenoCam	Validated against CMA stations (error <10 days)	Field observations

5. The reported missing rate and error rate are presented without explanation of how they were calculated. No analysis is provided on how different QC steps affect uncertainty in trend estimation. The authors should clarify the methodology for estimating error rates and add a sensitivity or uncertainty propagation analysis.

Response: We thank the reviewer for this important methodological suggestion. We have addressed both concerns in the revised manuscript.

Regarding the calculation methodology for missing rate and error rate: We have added explicit descriptions in Sections 2.3 and 5.4. The missing rate was calculated as the proportion of missing phenological records relative to the total expected observations (101 stations × 44 years × 10 phenological stages = 44,440 records). The error rate was calculated as the proportion of records corrected during the QC process relative to the total valid observations after imputation: error rate = (number of corrected records) / (total records − missing records) × 100%.

Regarding sensitivity analysis: We have added a new paragraph in Section 5.4 that compares trend estimates for silking (one of the most climate-sensitive stages) under three scenarios: (A) using the fully QC-corrected dataset; (B) excluding all stations with **fewer than 10 years of observations**; and (C) using uncorrected raw observations. The results demonstrate that the overall spatial patterns and the sign of trends remain consistent across all three scenarios, with trend magnitudes differing by **less than 10%** between scenarios. The largest differences were observed for stations with shorter observational records and for early-period records (1981–1995), where standard errors were notably larger, confirming that our QC procedures do not introduce directional bias but effectively reduce uncertainty in trend estimation.

We believe these additions provide transparent documentation of our quality metrics and demonstrate the robustness of our trend estimates to different QC treatments. All changes are highlighted in the revised manuscript. We thank the reviewer again for this thoughtful suggestion.

The revisions in the manuscript are reflected as follows:

Section 2.3: The missing rate was calculated as the proportion of missing phenological records relative to the total expected observations (101 stations × 44 years × 10 phenological stages = 44,440 records). A record was considered missing if no valid observation was recorded for a specific stage at a specific station in a specific year.

Section 5.4: The error rate was calculated as the proportion of records corrected during the QC process relative to the total valid observations after imputation (44,440 total records minus missing records). Specifically, the error rate = (number of corrected records) / (total records − missing records) × 100%. In total, 420 records were corrected through expert arbitration and cross-verification, yielding a final error rate of 0.96%.

To assess the sensitivity of our trend estimates to different QC treatments, we compared the decadal trends for silking (one of the most climate-sensitive stages) under three scenarios: (A) using the fully QC-corrected dataset; (B) excluding all stations with fewer than 10 years of observations; and (C) using uncorrected raw observations (i.e., before QC). The results show that the overall spatial patterns and the sign of trends remain consistent across all three scenarios, with trend magnitudes differing by less than 10% between scenarios. The largest differences were observed for stations with shorter observational records and for early-period records (1981–1995), where standard errors were notably larger, confirming that our QC procedures effectively reduce uncertainty in trend estimation without altering the fundamental direction of phenological shifts. We therefore recommend that users apply the fully QC-corrected dataset for robust trend analysis.

Minor to Moderate Comments:

1. No example code is provided for data reuse, and there is no guidance on which stations or years have lower data quality and should be used with caution.

Response： We thank the reviewer for these practical suggestions to improve data usability. We have addressed both concerns in the revised manuscript.

Regarding example code for data reuse: We have added a new **"Code Availability"** section in the revised manuscript, which states that a **Python script** for performing univariate linear regression analysis is available in the Science Data Bank repository alongside the dataset (https://doi.org/10.57760/sciencedb.32076). The script loads the core data tables, calculates linear trends for user-specified phenological stages or growth lengths, generates regression statistics (slope, intercept, R^2, p-value), and produces time-series plots with trend lines. The script is provided as a reference implementation and can be easily adapted for custom analyses using other software such as R, OriginPro, or Excel.

Regarding guidance on stations or years with lower data quality: We have addressed this concern through both enhanced documentation and transparent data provision, rather than imposing subjective quality labels. Specifically:

(1) We have added a **dedicated paragraph at the end of Section 5.4 (Assessment of Final Data Quality)** that explicitly advises users on data selection considerations. This paragraph recommends: (a) prioritizing stations with longer observational records for robust trend analysis; (b) exercising additional caution when interpreting early-period records (1981–1995), which derive from digitized paper archives; (c) considering the sparser station density in the southern HHHP (Jiangsu and Anhui provinces) when making regional-scale generalizations; and (d) focusing on the post-2000 period for applications requiring the highest data certainty.

(2) Rather than assigning simplified quality flags, our core data tables already include **station-specific summary statistics**—including the number of valid years, maximum, minimum, mean, standard deviation, and regression diagnostics (R^2, p-value, F-value)—for every phenological stage and growth length. These metrics allow users to **make their own informed judgments** about which stations or time periods are suitable for their specific research questions, without being constrained by a subjective classification imposed by us. This approach recognizes that different research applications may legitimately require different quality thresholds.

We believe this combination of explicit written guidance and rich metadata provides users with greater flexibility and transparency than a simplified quality labeling system. All revisions are highlighted in the revised manuscript. We appreciate the reviewer's thoughtful suggestions and trust that these enhancements sufficiently address the practical usability concerns raised.

2. Phrases such as unprecedented data quality are overly promotional and should be replaced with more neutral wording.

Response: We thank the reviewer for this suggestion. We have replaced all overly promotional language throughout the manuscript with more neutral wording. Specifically, we changed "unprecedented data quality" to "high data quality", "Unprecedented Data Quality and Scientific Value" to "Data Quality and Scientific Value", "unparalleled resource" to "valuable resource", and "unprecedented combination" to "distinctive combination". All revisions are highlighted blue in the revised manuscript.

We believe these revisions ensure the manuscript maintains a professional and appropriately neutral academic tone throughout.

3. References to existing phenology datasets are scarce, while self-citations are relatively frequent.

Response: We thank the reviewer for this constructive criticism regarding the balance and breadth of our literature review. We have carefully addressed both aspects of this concern in the revised manuscript.

(a) Regarding the scarcity of references to existing phenology datasets

We acknowledge that our original manuscript did not sufficiently cite recently published phenological datasets for maize in China. To remedy this, we have **added a dedicated paragraph in the Introduction** that systematically reviews the major existing datasets. This new paragraph reads as follows:

"In recent years, several phenological datasets for maize in China have been developed using remote sensing approaches. For example, Peng et al. (2024) produced the CCPD-Maize dataset, a 30-m resolution maize phenology product covering 22 provinces from 2001 to 2020, which achieved high agreement (R^2 = 0.964) with agrometeorological observations for emergence date. Other products include the 30-m maize phenology dataset by Niu et al. (2022) and the 1-km ChinaCropPhen1km dataset by Luo et al. (2020). Additionally, station-based observational datasets, such as the 2008–2022 winter wheat-summer maize phenology records from the Old Yellow River alluvial plain (Ding et al., 2024), provide valuable point-scale evidence. However, these existing products have inherent limitations: remote sensing products rely on proxy indicators (e.g., NDVI) rather than direct visual observations and may fail to capture the full sequence of ten key growth stages; existing station datasets cover shorter periods (typically <15 years) or limited spatial extent. Our dataset bridges these gaps by providing 44 years of direct visual observations across 101 stations, capturing all ten phenological stages from sowing to maturity."

We have also added the corresponding references to the reference list:

Peng, Q., Shen, R., Liu, Y., et al. (2024) Spatial and Temporal Patterns of Maize Phenology in China From 2001 to 2020. Journal of Geophysical Research: Biogeosciences, 129(10): e2024JG008095. https://doi.org/10.1029/2024JG008095

Niu, Q., Li, X., Huang, J., Huang, H., Huang, X., Su, W., & Yuan, W. (2022). A 30 m annual maize phenology dataset from 1985 to 2020 in China. Earth System Science Data, 14(6), 2851–2864. https://doi.org/10.5194/essd-14-2851-2022

Luo, Y., Zhang, Z., Chen, Y., Li, Z., and Tao, F. (2020) ChinaCropPhen1km: a high-resolution crop phenological dataset for three staple crops in China during 2000–2015 based on leaf area index (LAI) products, Earth Syst. Sci. Data, 12, 197–214. https://doi.org/10.5194/essd-12-197-2020

Ding, D. W., Yong, B. B., Ren, W., Xie, K., Zhao, Y. J., Wang, G. S., Chen, J. P., and Wang, M. H. (2024) The Dataset for Crop Phenology of Winter Wheat and Summer Maize in the Alluvium Plain of the Old Yellow River from 2008 to 2022, Journal of Agricultural Big Data, 6, 552–557 (in Chinese with English abstract)

In this new paragraph, we explicitly contrast the **limitations** of these existing products (remote sensing proxies, short temporal coverage, incomplete growth stages) with the **strengths** of our dataset (44-year span, direct visual observations, 101 stations, all ten growth stages). This comparison, we believe, better positions our dataset as a complementary and advanced resource rather than a redundant effort. We have also revised the transition sentence to ensure logical flow from the existing dataset review to the introduction of our work, now reading: "Building upon these existing efforts while addressing their limitations, the present study developed a comprehensive maize phenology dataset..."

(b) Regarding the frequency of self-citations

We have carefully reviewed our reference list to address this concern. Upon examination, we identified that our self-citations consist of two references, both of which serve essential scientific purposes rather than being redundant.

The first self-citation, Zhang et al. (2025a), refers to our team's previously published winter wheat phenology dataset for the same HHHP region, which employs an identical multi-tier quality control framework and represents a methodological precursor to the present study. Citing it provides essential methodological continuity and allows readers to understand the established QC protocols upon which this maize dataset builds. The second self-citation, Zhang et al. (2025b), is the data DOI citation for the dataset itself, which is required as part of the data availability statement and is standard practice in ESSD data papers. It enables direct findability and access to the dataset, aligning with the FAIR principles we advocate.

We have retained these self-citations as they serve legitimate scientific and documentation purposes. However, to further balance the reference composition, we have **substantially expanded our external literature coverage** by adding the four newly introduced maize phenology datasets (Peng et al., 2024; Niu et al., 2022; Luo et al., 2020; Ding et al., 2024) in the revised Introduction. This expansion significantly enriches the external citation base of our manuscript and appropriately reflects the broader field.

We are confident that these revisions substantially strengthen the literature review, clearly articulate the unique contributions of our dataset relative to existing products, and maintain a balanced and professionally appropriate citation composition. We again thank the reviewer for this thoughtful and helpful suggestion.

4. FAIR is misspelled in the abstract.

Response： We thank the reviewer for their meticulous reading of our manuscript. Upon careful re-examination, we confirm that the acronym **FAIR** and its full expansion in the abstract are **spelled correctly** following the original definition by Wilkinson et al. (2016):

"FAIR (Findable, Accessible, Interoperable, and Reusable)"

Specifically, in our abstract (Line 1), the text reads:

"The dataset presents a FAIR (Findable, Accessible, Interoperable, and Reusable), comprehensive, long-term dataset..."

We have verified all four constituent words against the standard definition and confirm that no typographical error exists. To avoid any potential ambiguity, we have ensured that **FAIR** appears uniformly in capital letters throughout the revised manuscript and that the full expansion is presented consistently.

We appreciate the reviewer's careful attention to detail and hope this clarification is satisfactory.

5. Figure 4 does not indicate trend significance.

Response： We thank the reviewer for this observation. In the revised manuscript, we have added the p-values directly to each of the time-series scatter plots in Figure 4. The p-value for each regression trend is now displayed adjacent to its corresponding regression line, allowing readers to directly assess the statistical significance of the observed trends.

6.Table 1 appears to be missing.

Response： We thank the reviewer for bringing this to our attention. However, we would like to respectfully clarify that **table (now Table 2 in the revised manuscript) was already included** and remains fully present in both the Word and PDF files submitted to the journal.

Specifically, **Table 2** is located in end of **Section 5.3** and is explicitly cited in the sentence: "These criteria, summarized in Table 1, ensure that an observation of 'silking' or 'maturity' represents the same plant developmental state..." (see original manuscript, Section 5.3).

Table 2 Standardized morphological criteria for identifying ten maize phenological stages at CMA agrometeorological stations

Phenology	**Observation basis**
Sowing (SO)	Actual sowing date.
Emergence (EM)	First leaf emerges from coleoptile, approximately 3.0 cm long.
Three-leaf (TS)	Third leaf emerges from second leaf sheath, approximately 2.0 cm long
Seven-leaf (SL)	Seventh leaf emerges from sixth leaf sheath, approximately 2.0 cm long.
Jointing (JO)	Basal internodes transition from flattened to round shape; near-ground stem nodes become round and firm to touch, approximately 3.0 cm long. Tassel differentiation begins at this stage.
Tasseling (TA)	Tassel tip emerges from leaf sheath.
Flowering (FL)	Anthers from upper-middle section of tassel become visible and release pollen.
Silking (SI)	Silks emerge from husks of female ear.
Milking (MI)	Silks darken to reddish-brown, outer husks fade slightly but remain green. Kernels reach normal size with thick white milky fluid in lower-middle section of ear.
Maturity (MA)	Over 80% of plants exhibit yellowed outer husks, dried silks, hardened kernels displaying variety-specific color. Kernels resist fingernail penetration.

We suspect that the reviewer may have experienced a **rendering issue** in their PDF viewer, as the table is formatted as a standard three-line table that should display correctly under normal conditions.

To prevent any confusion and further enhance clarity, we have made the following improvements in the revised manuscript:

- **Added an additional cross-reference** to Table 1 in Section 5.4, reinforcing its connection to data quality assurance.
- **Refined the table caption** to more explicitly describe its content (*"Standardized morphological criteria for identifying ten maize phenological stages at CMA agrometeorological stations"*), making it immediately clear even upon rapid browsing.

We appreciate the reviewer's careful reading and hope that these revisions will eliminate any possible ambiguity.

We sincerely thank the reviewer once again for the thorough and constructive evaluation of our manuscript. The insightful comments and detailed suggestions have been invaluable in substantially

improving the clarity, rigor, and overall quality of our work. We believe that the revised manuscript now presents the dataset and its validation with greater transparency and scientific robustness. Should the reviewer still find any aspect unsatisfactory, we remain fully committed to making further revisions as needed. We greatly appreciate the reviewer's time and effort in helping us strengthen this contribution to ESSD.